European Commission

JRC TECHNICAL REPORTS

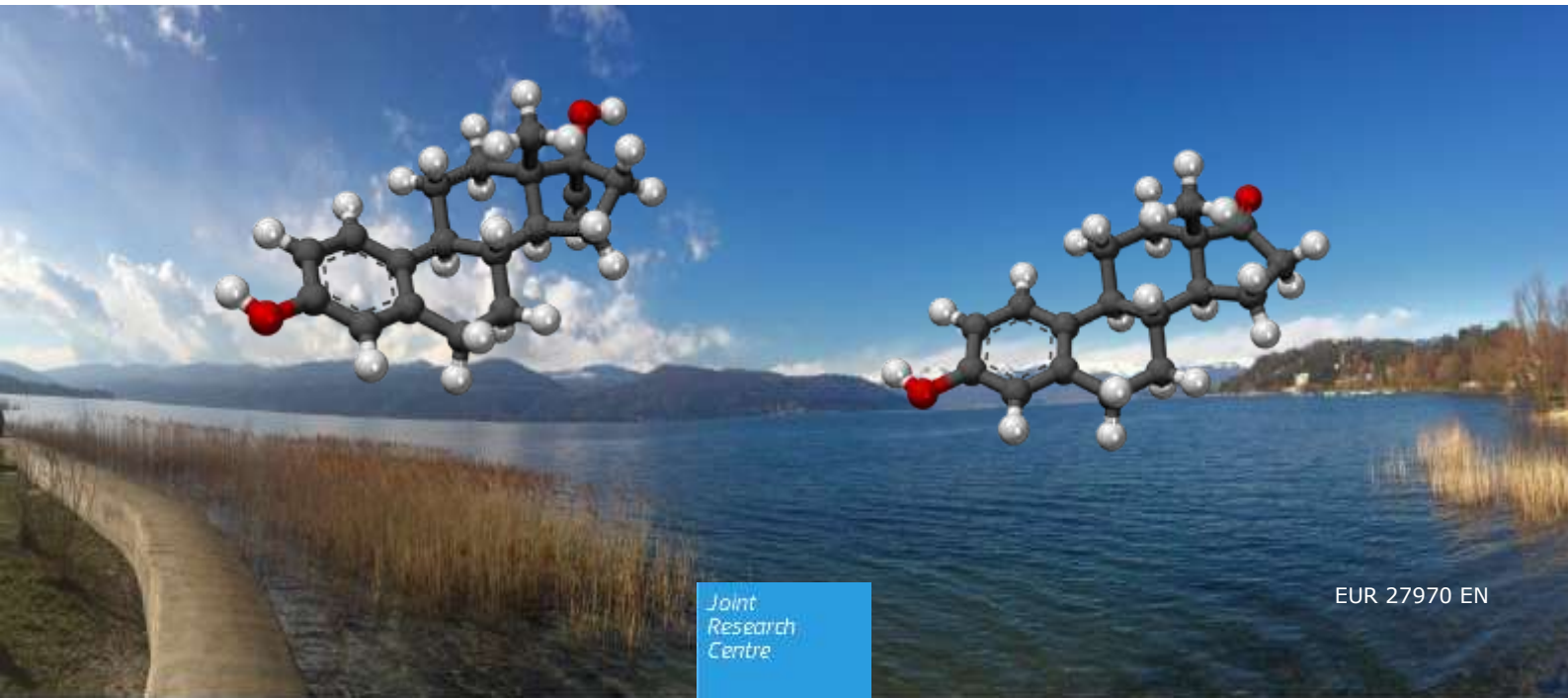
Water Framework Directive Watch List Method

Analysis of 17β-estradiol and estrone

Validation report, according to ISO 17025 requirements

S. Tavazzi, S. Comero, M. Ricci,
B. Paracchini, G. Mariani,
and B. M. Gawlik

2016



EUR 27970 EN

Contact information Simona Tavazzi
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 120, 21027 Ispra (VA), Italy
E-mail: simona.tavazzi@jrc.ec.europa.eu
Tel.: +39 0332 78 3683
Fax: +39 0332 78 9328

JRC Science Hub
https://ec.europa.eu/jrc

JRC94947

EUR 27970 EN

Print	ISBN 978-92-79-59358-1	ISSN 1018-5593	doi: 10.2788/46705
PDF	ISBN 978-92-79-59357-4	ISSN 1831-9424	doi: 10.2788/122189

Luxembourg: Publications Office of the European Union,

How to cite: Authors: S. Tavazzi, S. Comero, M. Ricci, B. Paracchini, G. Mariani, B. M. Gawlik; Water Framework Directive Watch List Method Analysis of 17β-estradiol and estrone, EUR 27970 EN; doi: 10.2788/122189

1 TABLE OF CONTENTS

Acknowledgements

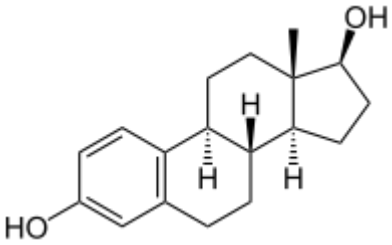
The authors gratefully acknowledge the support provided by Gráinne Mulhern (EC-JRC) for the revision of English language and improving the readability of the text.

Abstract

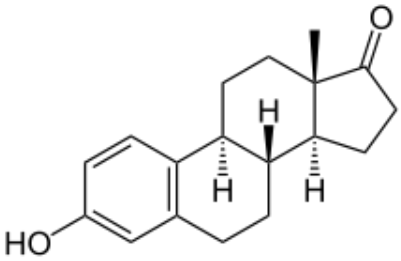
Validation of an analytical method is a necessary step in controlling the quality of quantitative analysis. Method validation is an established process which provides documentary evidence that a system fulfils its pre-defined specification, or shows that an analytical method is acceptable for its intended purpose. The purpose of the present study was to validate an SPE-LC-MS/MS method for the determination of 17b-estradiol and estrone in surface water samples according to the requirements laid down in ISO 17025. The calibration curves, working ranges, recoveries, detection and quantification limits, trueness as well as repeatability were determined. The uncertainty budget was estimated following a top-down approach based on in-house validation data. For 17β-estradiol, the expanded relative uncertainty was estimated as 16 and 15% at low and high concentration levels respectively, based on 15 replicate measurements on 5 days, and as 50 and 44% for a single measurement performed on a single day. For estrone, the expanded relative uncertainty was estimated as 13 and 12% at low and high concentration levels respectively, based on 15 replicate measurements on 5 days, and as 42 and 38% for a single measurement performed on a single day.

1 Introduction

17β-estradiol chemical structure:



Estrone chemical structure:



Chemical property data		
	17β-Estradiol	**Estrone**
Formula	$C_{18}H_{24}O_2$	$C_{18}H_{22}O_2$
Molecular mass	272.38g/mol	270.37 g/mol
CAS number	50-28-2	53-16-7
Systematic IUPAC name	(17β)-estra-1,3,5(10)-triene-3,17-diol	3-hydroxyestra-1(10),2,4-trien-17one

Steroid hormones are biologically active compounds; synthetised from cholesterol and secreted by the adrenal cortex, testes, ovaries and placenta in humans and animals.

Among them, estrogens, such as 17β-estradiol and estrone, are predominantly female hormones that are important for maintaining the health of reproductive tissues, breasts, skin and brain.

All humans as well as animals can excrete steroid hormones from their bodies, which end up in the environment through sewage discharge and animal waste disposal.

Their feminisation effects in invertebrates and fish have been confirmed worldwide (Caldwell *et al.*, 2008)

In the first review of the list of priority substances (PS) under the European Water Framework Directive (WFD), the Commission's legislative proposal introduced an environmental quality standard (EQS) for 17β-estradiol in the aquatic environment. The annual average EQS value (AA-EQS) for 17β-estradiol was set at $4x10^{-4}$ µg/l in inland surface waters (including rivers, lakes and related artificial or heavily modified water bodies), and at $8x10^{-5}$ µg/l in other (coastal) surface waters.

No AA-EQS has been established for estrone, for which the predicted no-effect concentration (i.e.: PNEC) in fresh water is known to be $3.6x10^{-3}$ µg/l.
Both compounds have been put on the newly introduced "watch list" mechanism to collect European Union-wide monitoring data (EU, 2013).

The analytical method reported in this report is based on solid-phase extraction (SPE) using a universal reversed-phase sorbent (HLB: hydrophilic-lipophilic balance) followed by ultra-high-pressure liquid chromatography coupled to tandem mass spectrometry (UHPLC-MS/MS) analysis. The procedure has been fully characterised in terms of linearity, working range, selectivity, matrix effect, precision, repeatability, trueness and uncertainty.

The objective of this method validation exercise is to contribute to the chemical analysis of natural estrogens in the aquatic environment in compliance with the first watch list exercise for the implementation of the Water Framework Directive (WFD).

2 Materials and methods

2.1 Chemicals

2.1.1 Standards

17β-Estradiol, CAS 50-28-2, batch 10407, purity 98.5%, expiry date May 23, 2015, Dr Ehrenstorfer, LCG Standard S.r.l., Milan, Italy

Estrone, CAS 53-16-7, batch SZE8123X, purity 99.3%, expiry date May 02, 2015, Sigma Aldrich, MO (USA).

17β-estradiol (3, 4 $^{13}C_2$, 99%), 100 µg/ml in acetonitrile, batch SDAB-007, purity > 98%, expiry date May 21, 20120, CIL, MA (USA).

2.1.2 Materials and reagents

Ethyl acetate for trace analysis (Carlo Erba Reactifs-SDS).

Methanol, code 701091.1612, (LC-MS) PAI, Panreac Quimica, Barcelona (Spain).

MilliQ water obtained from a MilliQ water system, Millipore, Bedford, MA (USA).

Ammonium hydroxide, 28% in water, 99.99% metals basis, code 338818, Sigma Aldrich, Germany.

OASIS HLB cartridges 6CC (0.2 g), code WAT106202, Waters, Milford, MA, USA.

2.1.3 Reagent solutions

Mobile phase A: 0.1% NH_4OH

a. Withdraw 0.385ml of ammonium hydroxide.
b. Dilute to 1-l volume with water.

Mobile phase B: methanol

a. Degas in an ultrasonic bath for 20 seconds.

UHPLC Autosampler strong washing solution

a. Transfer 900 ml of water and 100 ml of methanol into a 1-l bottle.
b. Mix and degas in an ultrasonic bath for 20 seconds.

UHPLC Autosampler weak washing solution

a. Transfer 100 ml of water and 900 ml of methanol into a 1-l bottle.
b. Mix and degas in an ultrasonic bath for 20 seconds.

UHPLC Seal washing solution

a. Transfer 100 ml of methanol and 900 ml of water into a 1-l bottle.
b. Mix and degas in an ultrasonic bath for 20 seconds.

UHPLC-MS/MS Reconstituting solution for LC-MS/MS analysis

a. Transfer 900 ml of mobile phase A into a 1-l bottle.
b. Add 100 ml of mobile phase B and mix.

2.1.4 Standard solutions

17β-estradiol stock standard solution (1 100 µg/ml)

a. Accurately weigh approximately 11.1 mg of 17β-estradiol in a 10-ml volumetric flask.
b. Dissolve and dilute to volume with methanol and mix.

Estrone stock standard solution (1 004 µg/ml)

a. Accurately weigh approximately 10.04 mg of estrone in a 10-ml volumetric flask.
b. Dissolve and dilute to volume with methanol and mix.

17β-estradiol and estrone working standard solution A (1 110 ng/ml, 1 004 ng/ml)

a. Withdraw 0.05 ml of 17β-estradiol stock standard solution and 0.05 ml of estrone stock standard solution into a 50-ml volumetric flask.
b. Dissolve and dilute to volume with methanol and mix.

17β-estradiol and estrone working standard solution B (1.11 and 1.004 ng/ml)

a. Withdraw 0.05 ml of 17β-estradiol working standard solution A and 0.05 ml of estrone working standard solution A into a 50-ml volumetric flask.
b. Dissolve and dilute to volume with methanol and mix.

Standard Solution A (17β-estradiol 0.111 ng/ml; estrone 0.1004 ng/ml)

a. Transfer 100 µl of working standard solution B into a 1-ml dark vial.
b. Dilute to 1 ml with methanol and mix.

Standard Solution B (17β-estradiol 0.555 ng/ml; estrone 0.502 ng/ml)

a. Transfer 500 µl of working standard solution B into a 1-ml dark vial.
b. Dilute to 1 ml with methanol and mix.

Standard Solution C (17β-estradiol 11.1 ng/ml; estrone 10.04 ng/ml)

a. Transfer 10 µl of working standard solution A into a 1-ml dark vial.
b. Dilute to 1 ml with methanol and mix.

Standard Solution D (17β-estradiol 55.5 ng/ml; estrone 50.2 ng/ml)

a. Transfer 50 µl of working standard solution A into a 1-ml dark vial.
b. Dilute to 1 ml with methanol and mix.

Standard Solution E (17β-estradiol 111 ng/ml; estrone 100.4 ng/ml)

a. Transfer 100 µl of working standard solution A into a 1-ml dark vial.
b. Dilute to 1 ml with methanol and mix.

Standard Solution Low QC (17β-estradiol 0.333 ng/ml; estrone 0.301 ng/ml)

a. Transfer 300 µl of working standard solution B into a 1-ml dark vial.
b. Dilute to 1 ml with methanol and mix.

Standard Solution High QC (17β-estradiol 99.9 ng/ml; estrone 90.4 ng/ml)

c. Transfer 90 µl of working standard solution A into a 1-ml dark vial.
d. Dilute to 1 ml with methanol and mix.

2.1.5 Internal standard solutions

Internal standard stock solution

17β-estradiol $^{13}C_2$, 100 µg/ml in acetonitrile

Internal standard working solution 1 (17β-estradiol $^{13}C_2$, 5 µg/ml)

a. Transfer 250 µl of 17β-estradiol $^{13}C_2$ stock solution into a 5-ml volumetric flask.
b. Dilute to 5 ml with methanol and mix.

2.2 Apparatus

2.2.1 Laboratory equipment

Analytical balance:	Model AX204, Mettler-Toledo SpA
Automatic pipettes:	Eppendorf Research (Milan, Italy)

Microsyringes: Microliter Syringes, Hamilton (Reno, CA, USA)

Autosampler vials for LC-MS: Micro-V vials target Dp clear, 1.5 ml, 12x22 mm National Scientific (Germany)

Volumetric flasks: Grade A various sizes, Duran®

Volumetric pipettes: Grade A various sizes, Duran®

Dionex Autotrace AT280 automated SPE system (Thermo Scientific, Waltham, MA, USA)

TurboVap II (Caliper Life Science, Mountain View, CA, USA)

Vortex Genius, Ika, Staufen, Germany

2.2.2 Instrumental equipment and conditions

2.2.2.1 UHPLC equipment and conditions

Pumps: Binary Solvent Manager, Model UPB, Waters (Milford, MA, USA)

Autosampler: Sample Manager, Model UPA, Waters (Milford, MA, USA)

Detector: QTRAP 5500, Applied Biosystems MDS SCIEX, (Foster City, CA, USA) equipped with Turbo V™ ion source

Flow rate: 300 µl/min

Injection volume: 5 µl

Analytical column: BEH C18, 1.7 µm, 50 x 2.1 mm, Waters, (Milford, MA, USA)

Mobile phase A: 0.1% ammonium hydroxide

Mobile phase B: methanol

The chromatography was performed in gradient mode according to the following scheme:

Time	A	B	Flow (ml/min)
0	90	10	0.3
0.5	80	20	0.3
1	60	40	0.3
5	20	80	0.3
6	20	80	0.3
6.5	90	10	0.3
8	90	10	0.3

Under these conditions, 17β-estradiol and estrone co-eluted at about 4.5 minutes. The run time was 8 minutes.

2.2.2.2 QTRAP 5500 operative condition

An AB Sciex QTRAP5500 mass spectrometer equipped with Turbo V™ ion source was used. The instrument was previously tuned and calibrated in electrospray mode using polypropylene glycol (PPG). Prior to analysis, all the specific parameters were optimised infusing a 1 µg/ml standard solution of analyte and internal standard (IS).

The eluent from the column was introduced directly into the ion source. The rapid desolvation and vaporisation of the droplets minimises thermal decomposition and preserves their molecular identity. The data were collected using the software programme Analyst 1.6.

All calculations were based on chromatographic peak area ratios for the multiple reaction monitoring (MRM) precursor-product ion transitions for analyte to the precursor-product ion transition of the internal standard. The general operating conditions were as follows:

Scan Type: Scheduled MRM

Polarity: Negative

Ion Source: Turbo Spray

Resolution Q1: Unit

Resolution Q3: Unit

MR Pause: 5.0000 msec

Analyte MRM	Time (min)	Declustering Potential (DP)	Collision Energy (CE)	
			271>145	271>143
17β-Estradiol (271>145; 143)	4.5	-83	-60	-78
			269>145	269>143
Estrone (269>145; 143)	4.5	-100	-53	-74-
17β-Estradiol $^{13}C_2$ (273>145)	4.5	-215	-71	

Curtain gas (CUR):	25
Collision gas (CAD):	Medium
Temperature (TEM):	550
Ion Transfer Voltage (IS):	-4 500.
Entrance Potential (EP):	10.00
Collision cell Exit Potential (CXP):	-11.00
Ion Source gas 1 (GS1) :	55
Ion Source gas 2 (GS2) :	45

3 Experimental set up of method validation

Different experiments were carried out for the characterisation of the developed procedure in terms of limit of detection and quantitation, linearity and working range, recovery, trueness, repeatability, intermediate precision and sample and extract stability. In our approach, a calibration curve and quality control samples (QCs) were freshly prepared in MilliQ water in triplicate for five different days. Some of the experiments were used in the evaluation of different parameters.

The analyte / internal standard peak area ratios will be used as target parameters for quantitation. A weighted (1/c) least-square regression analysis of data was performed to determine the calibration curve parameters and the coefficient of determination (R^2).

The equation obtained with the linear regression method is as follows:

$$X = \frac{Y - B}{A}$$

where:

X = analyte concentration

$$Y = \text{peak area ratio} = \frac{\text{analyte peak area}}{\text{I.S. peak area}}$$

A = slope

B = intercept.

Analyst 1.6 software was used for data acquisition and data processing.

Statistical calculations will be performed using Excel software.

3.1 Calibration curve

Calibration standards in MillliQ water (six different spiking levels, including a blank sample) covering the studied calibration range (0.11-111 ng/l for 17β-estradiol and 0.10-100.4 ng/l for estrone) were freshly prepared and processed on each day of validation.

The relationship (goodness of fit) between peak area ratios of analyte / I.S. and concentrations in the investigated concentration range was assessed by the coefficient of determination (R^2) and by the shape of the distribution of residuals around the horizontal axis.

The acceptance criteria set for calibration curves were:

- $R^2 \geq 0.9900$ calculated over five calibration curves and
- random dispersion of residuals around the horizontal axis, proving the pertinence of the linear regression model to interpret the data.

3.1.1 Matrix effect

A specific test for matrix effect was made by applying the standard addition method to typical test material.

Surface waters (i.e. Lake Maggiore, Ispra Bay, (VA), Italy) were sampled and spiked in a way that provides the same final dilution as the normal procedure produces in MilliQ water.

The range of addition encompassed the same range as the samples spiked in MilliQ water.

Analysis of covariance (ANCOVA, Field *et al.*, 2012) was first used to compare the calibration curves within each water type (i.e. MilliQ water and surface water) to check the stability over several days. Calibration curves were then compared within water types to assess whether a matrix effect exists.

ANCOVA was applied to the following three cases:

a. 5-day calibration curves for each compound in MilliQ water
b. 2-day calibration curve per each compound in surface water
c. 2 calibration curves per each compound measured in both MilliQ and surface water after establishing the stability over several days (cases a and b).

ANCOVA was performed using the R software (R Core Team, 2014) with the following variables specifications:

- *Std* the covariate variable = the concentration of the standard solution used to compute the calibration curve. Five concentration levels were used.
- *Computed*, the dependent variable = the computed concentration of the compound obtained from the peak area ratio.
- *Day,* the factor = the fixed factor which corresponds to the number of calibration days in cases a and b, and to the matrix type in case c.

An analysis of covariance was performed to establish whether all calibration curves obtained for each level of the factor have equal slopes and intercepts. This means verifying whether or not the factor has a significant effect on the dependent variable, "cleaned" by the effect of the covariate variable.

Depending on the case, the factor can have two or five levels. In case a, the five levels are given by the five different days on which the calibration curves are determined in MilliQ water. In case b, the two levels are given by the two different days on which the two calibration curves are determined in surface water. In case c, the two levels correspond to the calibration curves determined in both MilliQ and surface water after having verified the day-to-day stability of calibration curves in each water type separately.

3.2 Repeatability and intermediate precision

Quality control samples (QCs) were freshly prepared in MilliQ water and analysed in triplicate during validation days at two spiking levels (0.33 and 99.9 ng/l for 17β-estradiol and 0.3 and 90.4 ng/l for estrone) for a total of 15 independent sample preparations.

The acceptance criterion set for the relative standard deviation (RSD) of the repeatability and intermediate precision was 20% at both spiking levels.

3.3 Limits of detection and quantification

The limits of detection and quantification were estimated by analysing 12 and 15 different blank samples for 17β-estradiol and estrone, respectively.

The mean value of blank samples (b) and the RSD served for LOD and LOQ estimation, were calculated in accordance with the following equations:

$$LOD = b + 3SD$$

$$LOQ = b + 10SD.$$

3.4 Trueness

Trueness was assessed by analysing three different QCs at two spiking levels (see 3.2) on each day of validation, and by the application of the significance test (t-test) at the 95% confident level.

Considering that the QCs were spiked using aliquot withdrawals from the respective working standard solutions, the t test was applied on a total of five samples (as the average result of the triplicate quality controls analysed for each day of validation) and four degrees of freedom. The difference between the mean value of spiked MilliQ water samples and their nominal concentrations was evaluated according to the following formula:

$$t = \frac{(x-\mu)\sqrt{n}}{SD} \quad t = \frac{(x-\mu)\sqrt{n}}{SD}$$

where x is the mean value of n samples with standard deviation SD, and μ is the nominal concentration.

3.5 Recovery

Recovery was evaluated by extracting and analysing in triplicate 1-litre MilliQ water samples spiked, before extraction, with native analytes only. The internal standard was then added to the extracts at the end of the sample preparation with the aim of allowing for the estimation of analytes lost during processing.

The recovery was evaluated by comparing the ratios of analyte/IS in spiked samples to the same ratios obtained by analysing a standard solution containing native compounds and the labelled solution at the same concentration levels, not subject to any handling.

The spiking level was 10 ng/l for both 17β-estradiol and estrone.

3.6 Stability

3.6.1 Stability of extract

Three extracts of QCs at each spiking level were re-injected after 2, 7 and 14 days of storage at 4°C in the dark. In order to be considered stable, the concentrations of re-injected samples have to fall within about twice the standard deviations of the concentration of QCs used for the repeatability study.

4 Preparation of calibration standards and water samples

4.1 Calibration standards and Quality Control samples (QCs)

a. Fill 1-l glass bottle with 1 l MilliQ water.
b. Add 1 ml of working standard solution according to the following:

Standard solution	17β-Estradiol conc. ng/ml	17β-Estradiol conc. in water ng/l	Estrone conc. ng/ml	Estrone conc. in water ng/l	IS conc. in water (ng/l)
A	0.111	0.111	0.1004	0.1004	50
B	0.555	0.555	0.502	0.502	50
C	11.1	11.1	10.04	10.04	50
D	55.5	55.5	50.2	50.2	50
E	111	111	100.4	100.4	50
Low QC	0.333	0.333	0.301	0.301	50
High QC	99.9	99.9	90.4	90.4	50

4.2 Water sample extraction

a. Add 10 µl of 17β-estradiol $^{13}C_2$ working solution to 1 l water standard and QCs
b. Shake the water sample
c. Condition SPE OASIS HLB cartridge with 10 ml of ethyl acetate
d. Condition SPE cartridge with 10 ml of methanol
e. Condition SPE cartridge with 10 ml of water
f. Load water samples at 10 ml/minute
g. Dry the sorbent under nitrogen for 30 minutes
h. Elute the sample with 10 ml ethyl acetate at 5 ml/minute
i. Reconstitute with 0.5 ml of reconstituting solution for LC-MS/MS analysis.

5 Validation procedure and results

5.1 Selectivity

For the identification of 17β-estradiol and estrone, two multiple reaction monitoring (MRM) transitions between the precursor ion and the two most abundant fragment ions were monitored. The first was used for quantification purposes, whereas the second ("qualifier") was used to confirm the presence of the target compound in the sample. The quantified analyte was identified by comparing the retention time of the corresponding standard and the isotopic ratio between two ions recorded (±30%) in the standard and water samples.

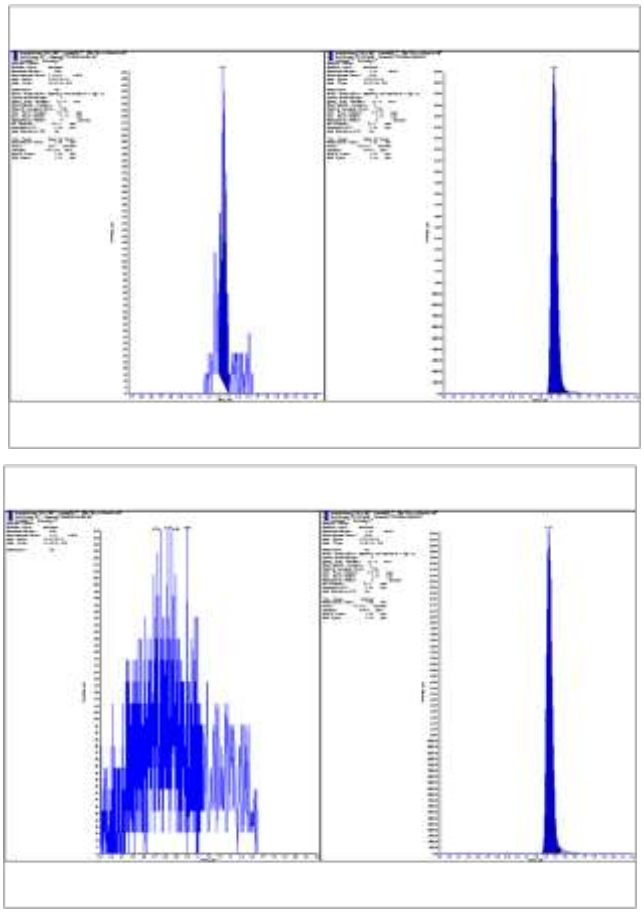
The selected mass transitions used for quantification and confirmation were respectively 271 > 145, 143 for 17β-estradiol, 269>145, 143 for estrone, and 271 > 145, 147 for 17β-estradiol $^{13}C_2$.

5.2 Limit of detection (LOD) and limit of quantification (LOQ)

Limits of detection and quantification were estimated by analysing blank samples.

A typical blank chromatogram is depicted in Figure 1. The left and right panels show instrumental signals for the quantification of MRM transitions of analyte and internal standard, respectively.

Figure 1 – Chromatograms for 17β-estradiol and estrone in blank sample



The mean values of the blank samples (b) and standard deviation (SD) were calculated using the data resulting from these experiments. The LOD and LOQ were estimated according to the formula reported in 3.3.

The results of the LOD and LOQ estimation are shown in **Error! Reference source not found.**.

Table 1 - LOD and LOQ values

Analyte	Nr of blanks analysed	LOD (ng/l)	LOQ (ng/l)
17β-Estradiol	12	0.03	0.08
Estrone	15	0.06	0.1

5.3 Linearity study

The linearity of the whole SPE-LC-MS-MS procedure was studied in the concentration range 0.11-111 ng/l for 17β-estradiol and in the range of 0.01-100.4 ng/l for estrone.

In order to verify the linearity of the calibration curve, a blank sample spiked only with labelled IS and five spiked MilliQ water samples (i.e. at 0.11, 0.55, 11.1, 55.5 and 111 ng/l for 17β-estradiol and at 0.10 0.5, 10.0, 50.2 and 100.4 ng/l for estrone) were extracted and analysed in three replicates on five different days. The calibration curves are illustrated in Figure 2a for 17β-estradiol and in Figure 2b for estrone.

Figure 2 – Calibration curves for 17β-estradiol

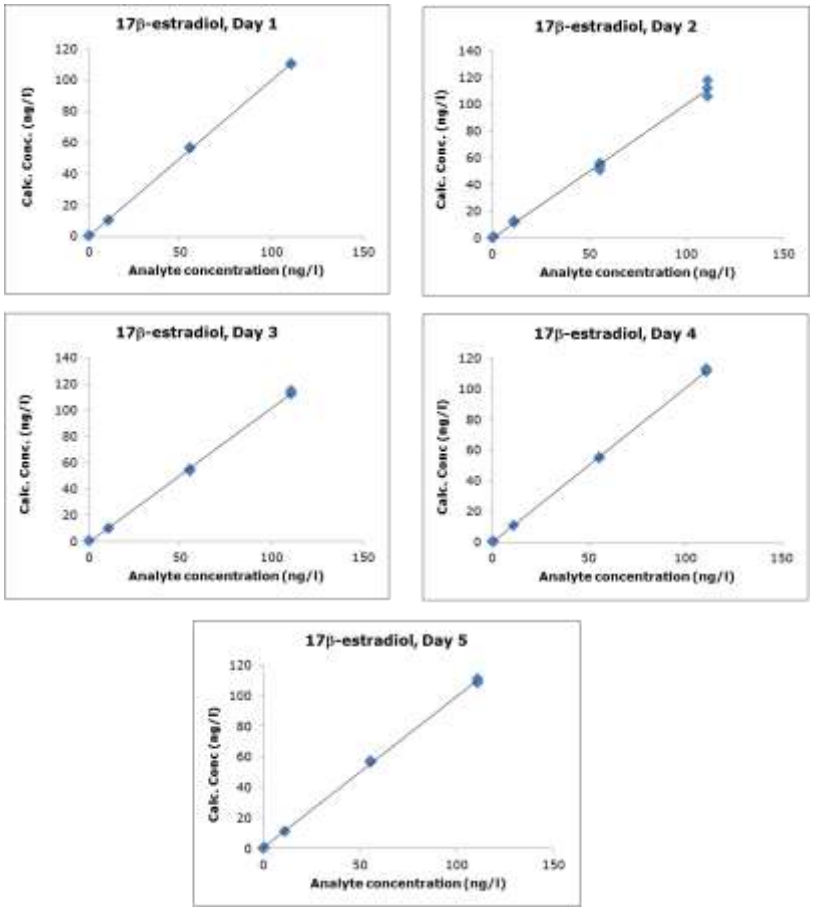
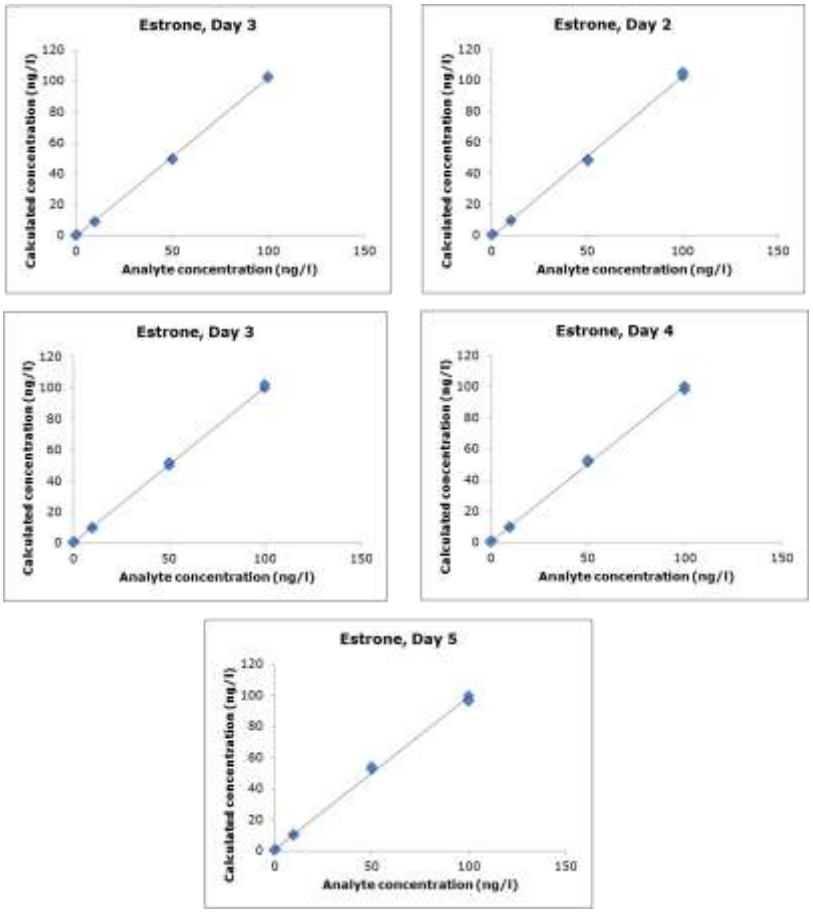


Figure 3 – Calibration curves for estrone



As reported in Table 2, the mean coefficient of determination (R^2) values, calculated over five calibration curves, were ≥0.998, with an RSD of 14.9% and ≥0.998, with an RSD of 8.0% for 17β-estradiol and estrone, respectively.

Table 2

Table 3 - Coefficient of determination (R^2) values for calibration curves on different days

Validation Day/Analyte	R^2	R^2
	17β-Estradiol	**Estrone**
1	0.9997	0.99963
2	0.9963	0.99834
3	0.9993	0.9994
4	0.9998	0.99928
5	0.9995	0.99775
Average	**0.9989**	**0.9989**
RSD (%)	**14.9**	**8.0**

The study of the distribution of residuals revealed shapes randomly dispersed around the horizontal axis, proving the pertinence of the linear regression model for interpreting the data. The residual plots are shown in Figure 4 and Figure 5.

Figure 4 – Residual plots for 17β-Estradiol

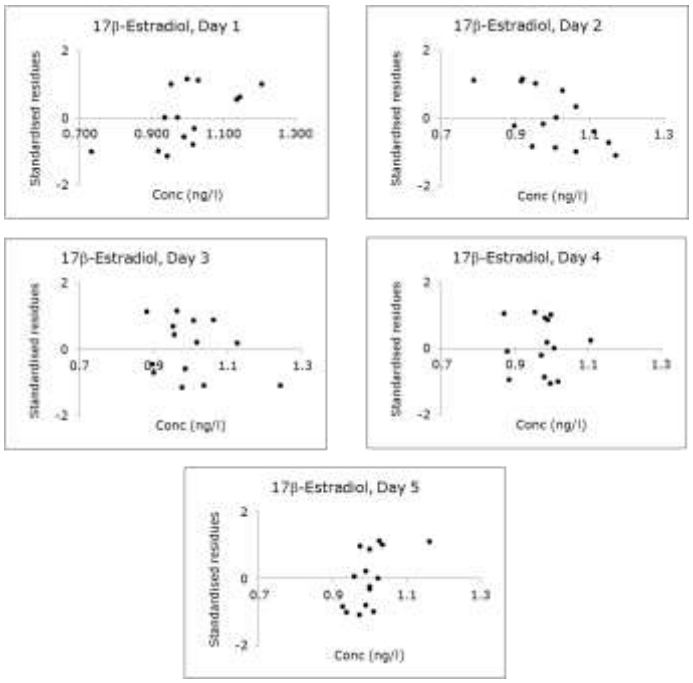
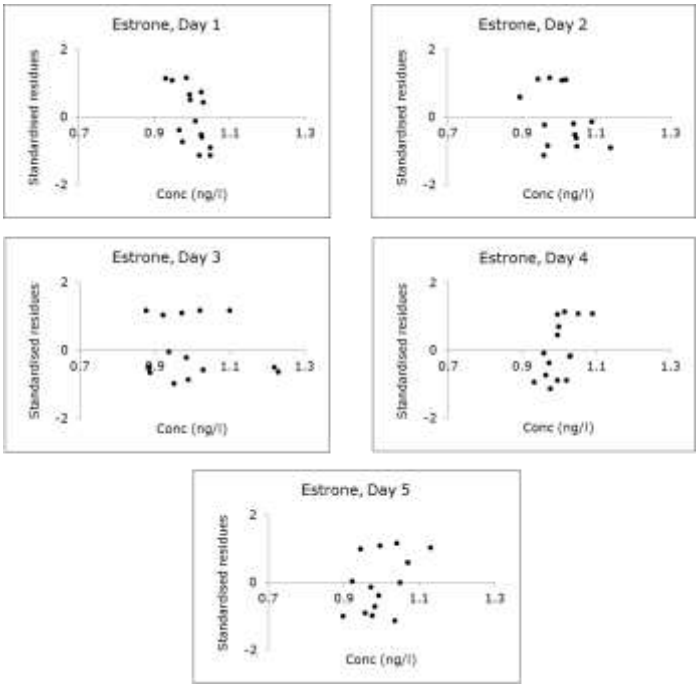


Figure 5 – Residual plots for Estrone



5.4 Matrix Effect

For a detailed statistical analysis of inputs and outputs, please refer to Annex 1.

A summary of the results is reported in the following sub-sections.

5.4.1 Verification of the ANCOVA assumption

Prior to the computation of the ANCOVA, the abovementioned assumptions must be verified for each of the two analysed compounds. All computations were carried out using R software (R Core Team, 2014).

5.4.1.1 Independence

This assumption implies that the covariate and any independent variables must be independent: it is verified by running an ANOVA with the covariate as the outcome and any independent variables (factors) as predictors to check that the covariate does not differ significantly across levels of these variables.

Since concentration levels of the standard solution (the covariate) were equal over time (cases a and b) and among matrices (case c), the p-value was 1 for all the cases and the hypothesis of independence was accepted.

5.4.1.2 Normality

This assumption could be checked by examining the residual plots from the fitted model for evidence of non-normality.

If the assumptions are satisfied, residuals should vary randomly around zero and the spread of the residuals should be about the same throughout the plot, with no systematic patterns. For the cases studied, the residuals did not suggest a time or matrix trend and the assumption of normality was accepted.

5.4.1.3 Homogeneity of variance

Levene's test was used to determine whether the variance in the outcome variable varies across groups.

For all of the tested cases, Levene's test results were non-significant, with p-values ranging between 0.9975 and 1. This means that the variances are very similar and the hypothesis of homogeneity of variances was accepted.

5.4.1.4 Linearity

The assumption of linearity was checked by looking at the individual plots of Y vs. X for each factor. No outliers should be identified.

5.4.1.5 Homogeneity of regression slopes

This assumption was verified by plotting a scatterplot for each experimental condition (factor) with the covariate on one axis and the outcome on the other.

The homogeneity of regression slopes is accepted if slopes are similar across groups.

From the scatterplots (plotted for the linearity assumption) it was clear that slopes are comparable.

5.4.2 ANCOVA Results

All statistical analyses were performed using R software (R Core Team, 2014).

5.4.2.1 Estrone: case a. – MilliQ water

The ANCOVA model was performed specifying five different slopes and five different intercepts (one per day). The model estimated 10 parameters from the data: five intercepts and five slopes.

Based on the output of the ANCOVA computation, the hypothesis of equal slopes (p-value>>0.05) and the hypothesis of equal intercepts (p-value=0.24) of regression lines were both accepted. It was therefore concluded that the day on which the calibration was computed did not influence the output variable (concentration of the analyte).

5.4.2.2 Estrone: case b. – Surface water

Here, the ANCOVA model was simpler, with only two different slopes and intercepts to compare. The model estimated four parameters from the data: two intercepts and two slopes.

From ANCOVA results, choosing a level of confidence of 95%, the hypothesis of equal slopes and intercepts between the regression lines was accepted (P-level = 0.07 for slopes, P-level = 0.93 for intercepts).

5.4.2.3 Estrone: case c. – Surface vs. MilliQ water

After testing the daily comparability of calibration curves in the MilliQ and surface water separately, it was possible to compare the calibration curves between the two matrices. In this case, the ANCOVA would give us information about the effect of the matrix type.

The calibration curve of the first day for each matrix type was taken for computatiing the ANCOVA.

Based on the results, the hypothesis of equal slope between the regression lines was accepted with a 95% confidence level (p-level=0.238). Moreover, the p-value for the hypothesis of equal intercepts between calibration curves was 0.802, indicating the acceptance of the null hypothesis.

The two calibration curves derived from the analysis of Estrone in surface and MilliQ waters can be assumed to be coincident at a level of confidence of 95%.

This implies that the matrix type has no significant effect on calibration curves for the considered analyte.

5.4.2.4 17β-estradiol: case a. MilliQ water

For case a, the hypothesis of equal slopes between the five calibration curves (p-value=0.378) and the hypothesis of equal intercepts (p-value=1.0) were both accepted based on the ANCOVA output,.

This indicated that there is no significant difference between the slopes and the intercepts of the calibration curves at a level of confidence of 95%.

5.4.2.5 17β-estradiol: case b. Surface water

For the interaction term (Day x Std), the probability value P higher is 0.942, indicating that there was no significant difference between the slopes of the calibration curves.

Moreover, the p-values for the effect of the day parameter was 0.867, indicating equal intercepts between calibration curves.

5.4.2.6 17β-estradiol: case c. Surface vs MilliQ water

After testing the daily comparability of calibration curves in the MilliQ and surface w aters separately, it was possible to compare the calibration curves between the two matrices.

Even in this case, there was no significant difference between the slopes of the calibration curves (p-value=0.69). Moreover, a p-value of 0.94 indicates equality of the intercepts between calibration curves. This implies that the matrix type had no significant effect on the calibration curves for the considered analyte.

5.5 Working Range

The working range, defined as the range of concentrations for which the chosen calibration curve is valid, was defined by the limits of quantification and highest point in the respective calibration curve. **Error! Reference source not found.** summarises the working range established for the procedure for 17β-estradiol and estrone, respectively.

Table 4 - Working range of the analytical method

Analyte	Working range (ng/l)
17β-Estradiol	0.08-111
Estrone	0.10-100

5.6 Trueness

Fifteen QCs at low and high concentration levels (i.e. about 0.3 and 90 ng/l) were extracted and analysed.

In order to demonstrate the truthfulness of the null hypothesis (H_0: the analytical method is not subject to systematic error), average back-calculated concentrations of three replicates per concentration level on five days of validation were considered and evaluated.

As reported in Tables 4 and 5, the t-values were found to be lower than the critical values for the target analytes at the studied concentration levels, demonstrating the absence of evidence of systematic errors in the quantification of the analytes.

Table 5 - Results of the trueness study for 17β-estradiol at different concentration levels

Analyte	Mean value (x) ng/l	Nr of samples (n)	Nr of degrees of freedom	Theoretical value (μ) ng/l	STD of samples (s) ng/l	Calculated t-value	Critical t_4 P=0.05	Decision
17β-Estradiol	0.34	5	4	0.33	0.04	0.57	2.78	OK
	105.74	5	4	99.9	11.7	1.12	2.78	OK

Table 6 - Results of the trueness study for estrone at different concentration levels

Analyte	Mean value (x) ng/l	Nr of samples (n)	Nr of degrees of freedom	Theoretical value (μ) ng/l	STD of samples (s) ng/l	Calculated t-value	Critical t_4 P=0.05	Decision
Estrone	0.33	5	4	0.301	0.035	1.85	2.78	OK
	98.5	5	4	90.4	9.4	1.93	2.78	OK

5.7 Recovery

The results of the recovery experiments, carried out according to section 3.5, are reported in Table 6.

Table 7 - Recovery

Analyte	Spike Level (ng/l)	Mean Recovery (%) (n=3)	SD (ng/l)	RSD (%)
17β-Estradiol	10	67.5	6.8	10
Estrone	10	70.7	7.0	9.9

5.8 Repeatability and intermediate precision

For repeatability and intermediate precision, QCs at two concentration levels were tested on five different days. For each sample, three replicate injections were made. Using a one-way ANOVA, the results were obtained as shown in Table 7 and Table 8 for 17β-estradiol and estrone, respectively.

Table 8 – RSDs of repeatability and intermediate precision for 17β-estradiol

Concentration level	Relative standard deviation (RSD_{Rep}) of Repeatability measurements	Relative standard deviation (RSD_{Ip}) of Intermediate precision measurements
Low	**21.7**	12.2
High	19.1	10.9

Table 9 – RSDs of repeatability and intermediate precision for estrone

Concentration level	Relative standard deviation (RSD_{Rep}) of Repeatability measurements	Relative standard deviation (RSD_{Ip}) of Intermediate precision measurements
Low	18.2	10.0
High	16.5	9.4

5.9 Stability of extract

Graphical representations of the stability of extracts are given in Figure 6 and Figure 7.

Since the concentrations evaluated by re-injections after storage for 2, 7 and 14 days at 4°C in the dark fall within about twice the standard deviations of the concentration of QCs used for repeatability evaluation, the extracts could be considered as being stable in the studied conditions.

Figure 6 – 17β-Estradiol: stability of extract after 2, 7 and 14 days storage at 4°C in the dark

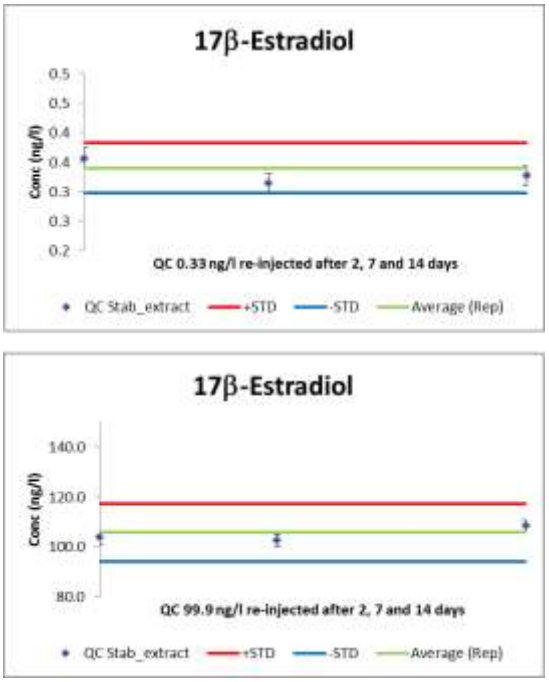
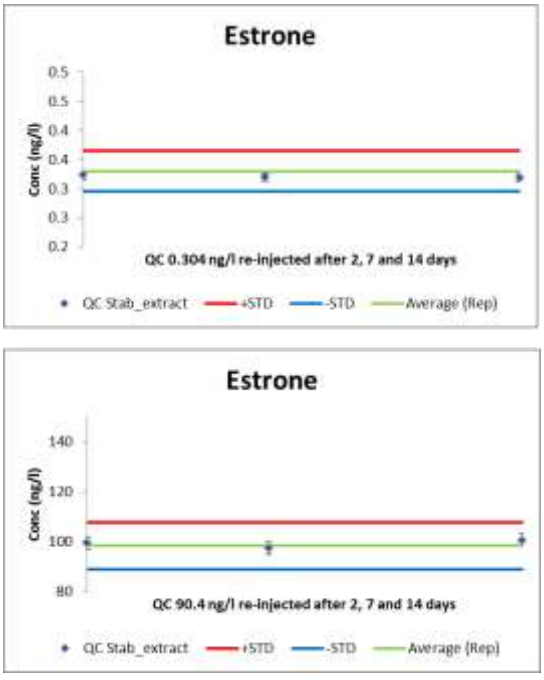


*Figure 7 – Estrone: s*tability of extract after 2, 7 and 14 days storage at 4°C in the dark



5.10 Uncertainty estimation

The estimation of measurement uncertainty was carried out following a top-down approach based on in-house validation data. The data derived from the validation of the method include the sample preparation, standard dilution, and chromatographic and mass spectrometric detection variability. This approach takes account of the relative standard deviation of repeatability, intermediate precision and trueness measurements. The uncertainty of prepared standard stock solutions is also considered, as another source of uncertainty.

The expanded uncertainty was calculated using the following formula: $U=k\sqrt{(u_{Tness})^2+(u_{Rep})^2+(u_{ip})^2+(u_{Std})^2}$

$$U = k\sqrt{(u_{Tness})^2 + (u_{Rep})^2 + (u_{Ip})^2 + (u_{Std})^2}$$

where:

U is the expanded relative uncertainty,

k is the coverage factor (k=2),

$u_{Tness}u_{Tness}$ is the relative standard uncertainty of trueness estimation,

u_{Rep} is the relative standard uncertainty of repeatability,

$u_{Ip}u_{Ip}u_{Ip}$ is the relative standard uncertainty of intermediate precision, and

u_{Std} is the relative standard uncertainty related to calibration standards including weighing, purity and dilution contributions.

5.10.1 Uncertainty of trueness

u_{Tness} is the standard relative uncertainty associated with the trueness, and has been calculated from the standard deviation (SD) of the mean of QCs used for trueness as follows:

$$u_{Tness} = \sqrt{\left(\frac{SD}{c\sqrt{n}}\right)^2}$$

where:

C is the average result of the QCs analyses,

n is the number of QCs that have been analysed.

5.10.2 Uncertainty of repeatability and intermediate precision

$u_{Rep}u_{Rep}$ and u_{Ip} are the standard relative uncertainties related to repeatability and intermediate precision measurements, respectively. Individual contributions are calculated according to the following equations:

$$u_{Rep}=\sqrt{\frac{(RSD_{Rep})^2}{n_{Rep}}}\quad u_{Rep} = \sqrt{\frac{(RSD_{Rep})^2}{n_{Rep}}}$$

and

$$u_{Ip} = \sqrt{\frac{(RSD_{Ip})^2}{n_{days}}} u_{Ip} = \sqrt{\frac{(RSD_{Ip})^2}{n_{days}}}$$

where:

RSD_{Rep} standard deviation of repeatability measurements

RSD_{Ip} standard deviation of intermediate precision measurements

n_{Rep} number of total replicates for repeatability measurements

n_{days} number of days for intermediate precision measurements.

5.10.3 Uncertainty of standard

u_{Std} is the standard relative uncertainty associated with 17β-estradiol, estrone standards used, and is calculated as follows:

$$u_{Std} = \sqrt{(u_{17\beta\text{-estradiol}})^2 + (u_{Flask})^2 + 2(u_{Syringe})^2 + (u_{Balance})^2}$$

in case of 17β-estradiol and as follows

$$u_{Std} = \sqrt{(u_{Estrone})^2 + (u_{Flask})^2 + 2(u_{Syringe})^2 + (u_{Balance})^2}$$

in the case of estrone, where:

For 17β-Estradiol, the Manufacturer's Certificate of Analysis reports the expanded combined uncertainty $\mathbf{U_{17\beta\text{-Estradiol}}}$ at 95% confidence level to be equal to 0.5%.

Consequently:

$$\mathbf{u_{17\beta\text{-Estradiol}}} = \mathbf{U_{17\beta\text{-Estradiol}}}/2 = 0.3.$$

For Estrone, the Manufacturer's Certificate of Analysis reports the expanded combined uncertainty $U_{\mathbf{Estrone}}$ for a single measurement at 95% confidence level with n=6 to be equal to 0.5%.

Consequently, the relative uncertainty is 0.2, according to the following formula:

$$U_{\mathbf{Estrone}} = U_{\mathbf{Estrone}}/2.45 = 0.2.$$

$\mathbf{u_{Flask}}$ is the uncertainty related to the volumetric flask. The tolerance of the class A 10-ml volumetric flask (given by the manufacturer) is set to 0.04 ml. As this value is not correlated with a confidence level or distribution information, a rectangular distribution is assumed. For the uncertainty estimation, the relative tolerance value (i.e. 0.4%) must by divided by $\sqrt{3}$ √3, giving a value of 0.231 for u_{Flask}.

$\mathbf{u_{Syringe}}$ is the uncertainty related to the withdrawal of the standard solution using a 1 000-µl Hamilton syringe. As these syringes are manufactured to be accurate within ± 1% of the nominal value and this value is not correlated with a confidence level or distribution information, a rectangular distribution is assumed. For the uncertainty estimation, the relative uncertainty (i.e. 1 ml/1 000 ml*100=0.1%) must by divided by$\sqrt{3}$ $\sqrt{3}$, giving a value for $u_{Syringe}$ equal to 0.058.

$\mathbf{u_{Balance}}$ is the contribution from the weight of standards, and it is due to the linearity uncertainty of the balance from the Calibration Certificate. From balance linearity (± 0.03 mg), a rectangular distribution is assumed to obtain a standard uncertainty; this contribution is considered twice, once for the tare and once for the gross weight. According to this approach, the $u_{Balance}$ as RSD % is:

$$u_{Balance} = 2x\sqrt{\left(\frac{0.03}{\sqrt{3}}\right)^2} = 0.035$$

$$u_{Balance} = \frac{0.035mg}{10mg}\% = 0.35\%.$$

As the repeatability and trueness of the measurement were estimated for two different concentration levels, the uncertainty can also be estimated separately for low and high concentration levels.

5.11 Final uncertainty budget

Considering the data for 17β-Estradiol, the uncertainty contribution from the repeatability assessment was 5.6% and 4.9% at low and high concentration levels respectively. The uncertainty contribution from intermediate precision was 5.5% and 4.9% at low and high concentration levels respectively. The uncertainty contribution from standard preparation was around 1%.

The expanded relative uncertainty was estimated a 16% and 15% at low and high concentration levels respectively, based on 15 replicate measurements on 5 days, and as 50% and 44% for a single measurement performed on a single day.

Considering the data for Estrone, the uncertainty contribution from the repeatability assessment was 4.7% and 4.3% at low and high concentration levels respectively. The uncertainty contribution from intermediate precision was 4.5% and 4.2% at low and high concentration levels respectively. The uncertainty contribution from standard preparation was around 1%.

The expanded relative uncertainty was estimated as 13% and 12% at low and high concentration levels respectively, based on 15 replicate measurements on 5 days, and as 42% and 38% for a single measurement performed on a single day.

The detailed uncertainty budget and results of the uncertainty estimations are reported in Table 9.

Table 10 - Uncertainty budget and estimated uncertainty of measurements

Estimated uncertainty	Values			
	17β-Estradiol		Estrone	
	Low concentration level (0.301 ng/l)	High concentration level (90.4 ng/l)	Low concentration level (0.333 ng/l)	High concentration level (99.9 ng/l)
u_{Tness} (%)	0.1	2.1	0.1	1..5
u_{Rep} (%)	5.6	4.9	4.7	4.3
u_{Ip} (%)	5.5	4.9	4.5	4.2
u_{Std} (%)	0.6	0.6	0.4	0.4
Expanded Relative Uncertainty (U, %) (k=2), (n_1=15, n_2=5)	**16**	**15**	**13**	**12**
Expanded Relative Uncertainty (U, %) (k=2), (n_1=1, n_2=1)	**50**	**44**	**42**	**38**

6 References

Caldwell DJ, Mastrocco F, Hutchinson TH, Linge R, Heijerick D, Janssen C, Anderson PD, Sumpter JP. (2008). Derivation of an aquatic predicted no-effect concentration for the synthetic hormone, 17α-ethinyl estradiol. Environmental Science & Technology, 42:7046-7054.

Field, A.; Miles, J.; Field Z.; Discovering Statistics Using R. SAGE Publications Ltd, 2012.

European Union (2013). "Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy".

Eurachem Group (1998). The Fitness of Purpose of Analytical Methods. A Laboratory Guid to Method Validation and Related Topics [Report].

Kidd KA, Blanchfield PJ, Mills KH, Palace VP, Evans RE, Lazorchak JM, Flick RW (2007). Collapse of a fish population after exposure to a synthetic estrogen. PNAS, 104 (21): 8897-8901.

Miller James N and Miller Jane C (2005). Statistics and Chemometrics for Analytical Chemistry, Fifth edition [Book]. .

Oaks JL, Gilbert M, Virani MZ, Watson RT, Meteyer CU, Rideout B, Shivaprasad HL, Ahmed S, Jamshed Iqbal Chaudhry M, Arshad M, Mahmood S, Ali A, Khan AA. (2004). Diclofenac residues as the cause of vulture population decline in Pakistan. Nature, 427:630-633.

Risebrough R. (2004). Fatal medicine for vultures. Nature, 427: 596-598.

R Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org (link verified on 02/06/2015)

González, F.J.E.; Hernández Torres, M.E.; Garrido Frenich, A.; Vidal J.L.M.; García Campana A.M.; Internal quality-control and laboratory-management tools for enhancing the stability of results in pesticide multi-residue analytical methods. Trends in Analytical Chemistry, 2004, 23, 5.

de Pinho, G.P.; Silvério, F.O.; Neves, A.A.; Lopes Ribeiro de Queiroz, M.E.; Evaluation of pesticide adsorption in gas chromatographic injector and column. Química Nova, 2012, 35, 4, p. 738-742.

Wang, L.; He, L.; Wang, Z.; Wang, X.; Shu, J.; Yang, J.; Zhang, G.; Zeng, Z.; Selection of a representative matrix for the multiresidue analysis of nine β-agonists in animal tissues and urine with LC-MS/MS. Analyst, 2013, 138, 4579.

List of Abbreviations and Symbols

Throughout this report, the following abbreviations and symbols are used:

ANOVA	Analysis of Variance	JRC	Joint Research Centre
ANCOVA	Analysis of Covariance	LC	Liquid chromatography
CAD	Collision Gas	LOD	Limit of detection
CUR	Curtain Gas	LOQ	Limit of quantification
CRM	Certified reference material	MRM	Multiple reaction monitoring
CXP	Collision Cell Exit Potential	MS	Mass spectrometry
DG	Directorate-General	PNEC	Predicted no effect concentration
EC	European Commission	PPG	Polypropylene glycol
EP	Entrance Potential	PS	Priority substances
EU	European Union	QC	Quality control sample
GS1	Ion Source gas 1	R^2	Coefficient of determination
GS2	Ion Source gas 2	RSD	Relative standard deviation
HLB	Hydrophilic-lipophilic balance	SD	Standard deviation
IES	Institute for Environment and Sustainability	SPE	Solid-phase extraction
		TEM	Temperature
IS	Internal standard/Ion Transfer voltage	UHPLC	Ultra-high-pressure liquid chromatography
ISO	International Organization for Standardization	WFD	Water Framework Directive

Chemical elements are identified by their respective symbols as defined by the International Union of Pure and Applied Chemistry (IUPAC)

List of Figures

List of Tables

Annex 1

ANCOVA test for the evaluation of the matrix effect in Estrone and 17β-Estradiol LC-MS/MS quantification in surface waters

2 TABLE OF CONTENTS

1. Introduction:

In the determination of Estrone and 17β–Estradiol in water samples, calibration curves prepared in MilliQ water were compared with those prepared in surface water. The comparison would indicate whether or not a significant matrix effect occurs for the selected analytes.

For this purpose, five calibration curves in MilliQ water and two calibration curves in surface water were determined on different days. Analysis of covariance (ANCOVA) was first used to compare the calibration curve within each water type to check the stability over several days. Calibrations were then compared between water types to assess whether a matrix effect exists.

The ANCOVA is a statistical tool that can be used to compare regression curves (slopes and intercepts). The ANCOVA is an extension of the analysis of variance (ANOVA) that provides a means of statistically controlling the (linear) effect of one or more continuous variables that are not part of the main experimental manipulation but have an influence on the dependent variable (Field *et al.*, 2012). These variables are called *covariates* and should be measured on an interval or ratio scale. A one-way ANCOVA evaluates whether population averages of the dependent variable are the same across all levels of a factor (independent variable), adjusting for differences in the covariate. The factor divides individuals into two or more groups or levels, while the covariate and the dependent variable differentiate individuals based on quantitative dimensions. The one-way ANCOVA is used to analyse data from several types of studies, including studies that investigate the differences among calibration curves in order to check their stability (González *et al.*, 2004), evaluate the existence of matrix effects (Wang *et al.*, 2013), and to compare different measurement procedures (de Pinho and Silvério, 2012).

2. Statistics

2.1. Assumptions

ANCOVA makes the same assumptions as ANOVA plus two considerations (points 1 and 5):

1. Independence: the covariate variable is independent of the groups (i.e. the covariant and independent variables are independent);
2. Normality: the residuals must be normally distributed around the regression line for each group;
3. Homogeneity of variance (homoscedasticity): the variance must be equal for both groups around their respective regression lines;
4. Linearity: the relationship between the dependent variable (y) and the covariate (x) is linear for each factor;
5. Homogeneity of regression slopes: the regression lines for these individual factors are assumed to be parallel (they have the same slope).

2.2. Problem under analysis

ANCOVA was applied in order to compare slopes and intercepts obtained for the following three cases:

a. Five-day calibration curves for compounds analysed in MilliQ water;
b. Two-day calibration curves for compounds measured in Surface water;

c. Two calibration curves for compounds measured in both MilliQ and Surface water after accepting the stability over several days (cases a and b).

All three cases were applied for the determination of Estrone and 17β-Estradiol.

ANCOVA was performed using the R software (R Core Team, 2014) with the following variables specifications:

- *Std*, the covariate variable = the concentration of the standard solution used to compute the calibration curve. Five concentration levels were users.
- *Computed*, the dependent variable = the computed concentration of the compound obtained from the peak area
- *Day*, the factor = the fixed factor which corresponds to the calibration day in cases a and b, and to the matrix type for case c.

The analysis of covariance was performed to establish whether, for each level of the factor, all calibration curves have equal slopes and intercepts. This means verifying whether or not the factor has a significant effect on the dependent variable, "cleaned" by the effect of the covariate variable.

Depending on the case, the factor can have two or five levels. In case a, the five levels are given by the five different days on which the calibration curves are determined in MilliQ water. In case b, the two levels are the two calibration curves determined in surface water. In case c, the two levels correspond to the calibration curves determined in both MilliQ and surface water after having verified the day-to-day stability of calibration curves in each water type separately.

Null hypotheses

The first null hypothesis of ANCOVA is that the slopes of the regression lines are all equal; in other words, the regression lines are parallel to each other. Once the null hypothesis that the regression lines are parallel is accepted, it is possible to test for the second null hypothesis: the intercepts of the regression lines are all the same.

Pre-analysis:

Five concentrations of the standard solution were analysed in three replicates. To conduct the ANCOVA analysis, the average values at each concentration level were computed and used.

3. Procedure

3.1. Verification of the ANOVA assumption

Prior to the computation of the ANCOVA, the abovementioned assumptions must be verified for each case of the two analysed compounds.

1. Independence

This assumption requires that the covariate and any independent variables are independent: it is verified by running an ANOVA with the covariate as the outcome and any independent variables (factors) as predictors to check that the covariate does not differ significantly across different levels of these variables. The R code, which is applied to each case and for the two compounds separately, is the following:

```
> independence<-aov(Std~Day, Edata)
> summary(independence)
```

The results are summarised in the following tables, where the R output is given separ
ately for each compound:

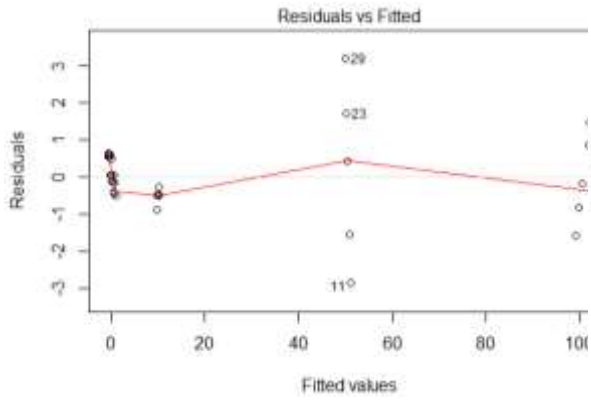
Case	Estrone: R output					
		Df	Sum	Sq Mean	Sq F	Value Pr(>F)
a	Day	4	0	0	0	1
	Residual	25	41554	1662		
b	Day	1	0	0	0	1
	Residual	10	16622	1662		
c	Day	1	0	0	0	1
	Residual	10	16622	1662		

Case	17β-Estradiol: R output					
		Df	Sum	Sq Mean	Sq F	Valiue Pr(>F)
a	Day	4	0	0	0	1
	Residual	25	51142	2046		
b	Day	1	0	0	0	1
	Residual	10	20457	2046		
c	Day	1	0	0	0	1
	Residual	10	20457	2046		

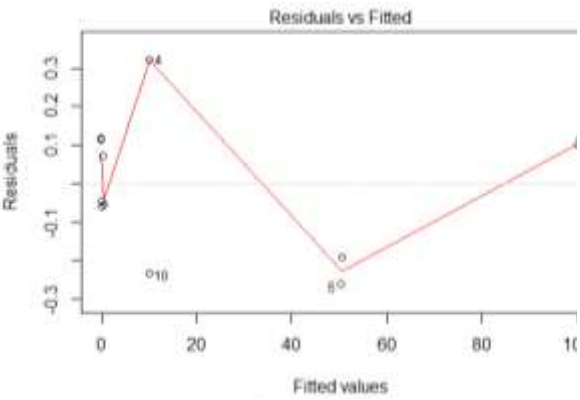
Since concentration levels of the standard solution (the covariate) are equal over the different timespans (cases a and b) and matrices (case c), the p-value is 1 for all the cases and the hypothesis of independence is accepted.

3.2. Normality

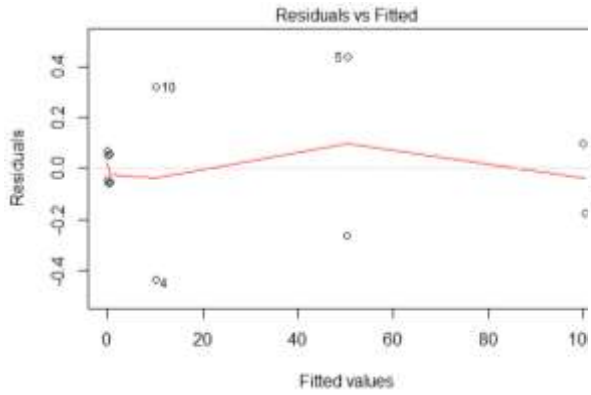
This assumption can be checked by examining the residual plots from the fitted model for evidence of non-normality.



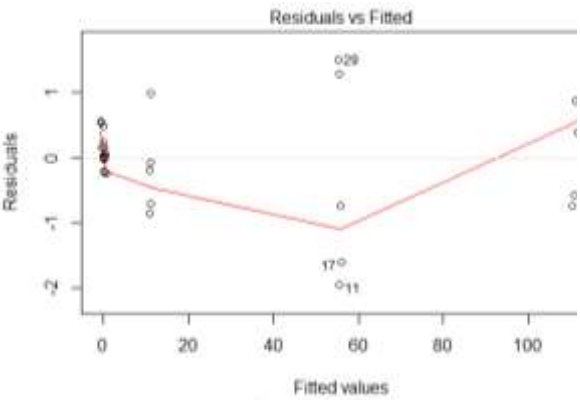
Estrone: case a



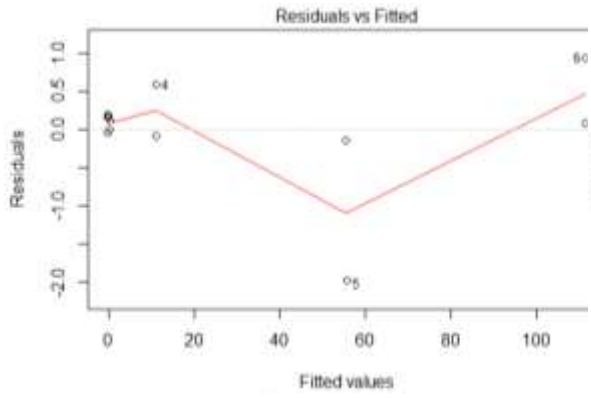
Estrone: case b



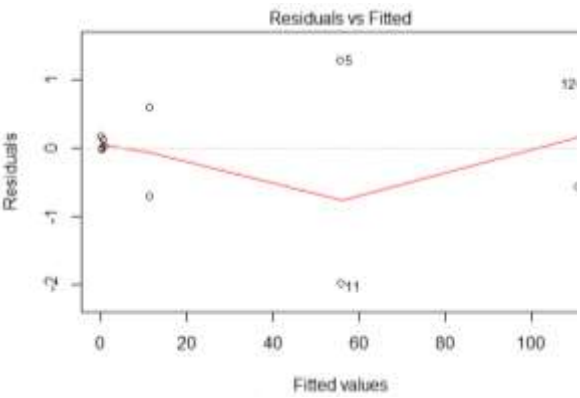
Estrone: case c



17β–Estradiol: case a



17β-Estradiol: case b



17β-Estradiol: case c

If the assumptions are met, residuals should vary randomly around zero and the spread of the residuals should be about the same throughout the plot, with no

systematic patterns. For these cases, the residuals do not suggest a time trend and the assumption of normality is accepted.

3.3. Homogeneity of variance

Levene's test is used to determine whether the variance in the outcome variable varies across groups. The R code is the following:

```
> leveneTest(Computed~Day, Edata)
```

R output of Levene's test is given in the following tables:

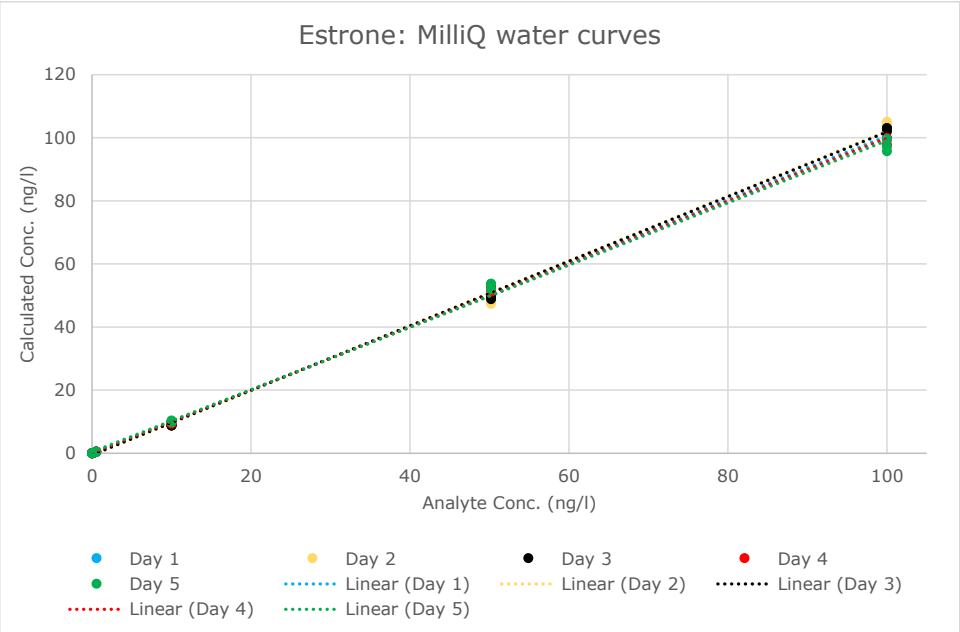
Case		Estrone: R output		
		Levene's test for Homogeneity of Variance (center = median)		
a		Df	F value	Pr(>F)
	group	4	0	1
	25			
b		Df	F value	Pr(>F)
	group	1	0	0.9994
	10			
c				
	Group			
	10	1	0	0.9988

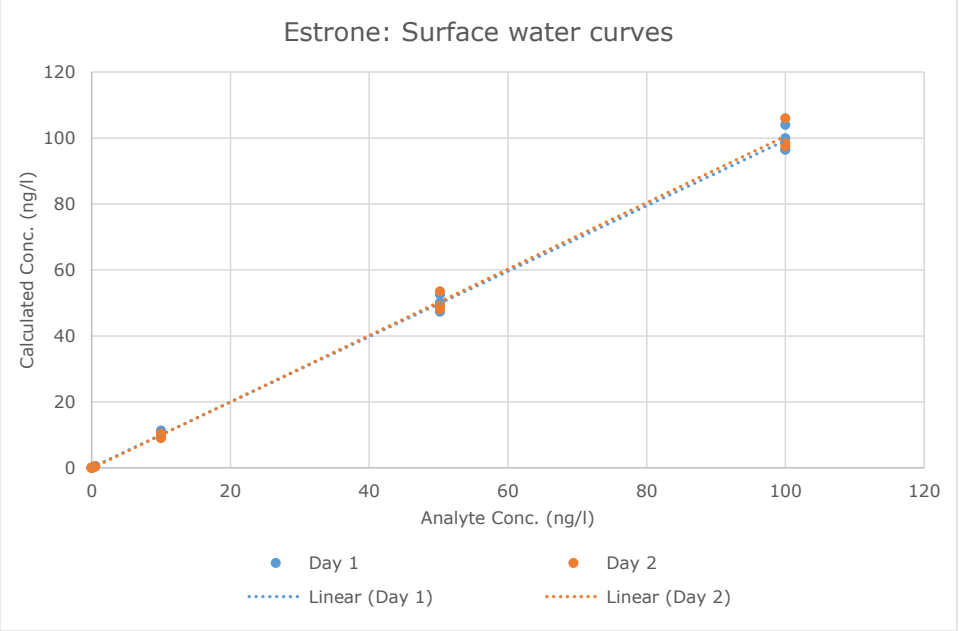
Case		17β-Estradiol: R output		
		Levene's test for Homogeneity of Variance (center = median)		
a		Df	F value	Pr(>F)
	group	4	0	1
	25			
b		Df	F value	Pr(>F)
	group	1	0	0.9975
	10			
c		Df	F value	Pr(>F)
	group	1	0	0.9996
	10			

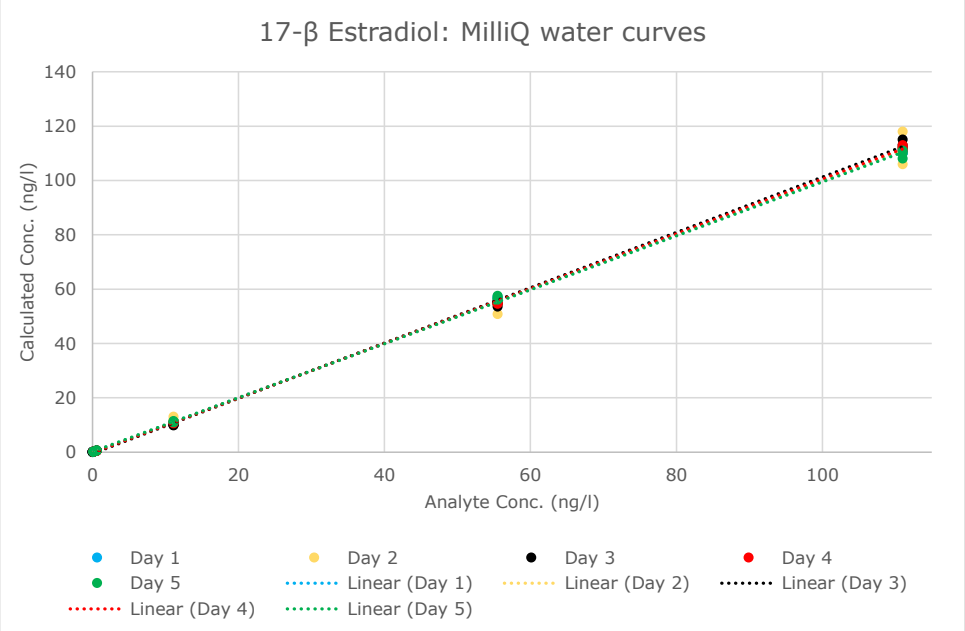
The output shows that Levene's test is non-significant for all cases, with p-values ranging between 0.9975 and 1. This means that the variances are very similar and the hypothesis of homogeneity of variances is accepted.

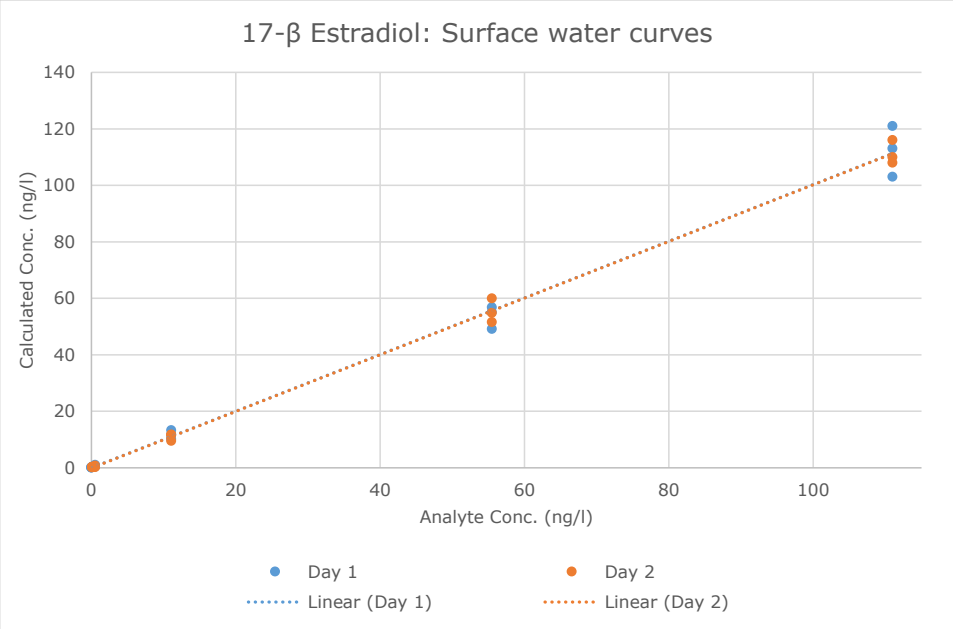
3.4. Linearity

The assumption of linearity is checked by looking at the individual plots of Y vs. X for each factor. No outliers should occur.


Estrone: MilliQ water curves
Calculated Conc. (ng/l)
Analyte Conc. (ng/l)
0
20
40
60
80
100
120
Day 1
Day 2
Day 3
Day 4
Day 5
Linear (Day 1)
Linear (Day 2)
Linear (Day 3)
Linear (Day 4)
Linear (Day 5)


Estrone: Surface water curves
Calculated Conc. (ng/l)
Analyte Conc. (ng/l)
0
20
40
60
80
100
120
Day 1
Day 2
Linear (Day 1)
Linear (Day 2)





3.5. Homogeneity of regression slopes

This assumption is verified plotting a scatterplot for each experimental condition (factor) with the covariate on one axis and the outcome on the other. The regression line for each of these scatterplots is then calculated, and the homogeneity of regression slopes is accepted if slopes are similar across groups. From the scatterplots (plotted for the linearity assumption) ,it is clear that slopes are comparable.

4. Results

4.1. Estrone: case a. – MilliQ water

In R, the ANCOVA model with five different slopes and five different intercepts (one per day) is specified using the following formatting:

```
> model_1<-lm(Computed~Day*Std, Edata)
> summary(model)
```

The R output is the following:

```
Call:
lm(formula = Computed ~ Day * Std, data = Edata)
```

Coefficients:				
	Estimate	Std. Error	T value	Pr(>|t|)
(Intercept)	-0.04917	0.65659	-0.075	0.941
DayDay2	-0.56464	0.92856	-0.608	0.550
DayDay3	-0.47042	0.92856	-0.507	0.618
DayDay4	0.19961	0.92856	0.215	0.832
DayDay5	0.50028	0.92856	0.539	0.596
Std	1.00558	0.01432	70.240	<2e-16 ***
DayDay2:Std	0.02242	0.02025	1.107	0.281
DayDay3:Std	0.01767	0.02025	0.873	0.393
DayDay4:Std	-0.00829	0.02025	-0.409	0.687
DayDay5:Std	-0.01943	0.02025	-0.960	0.349
Signif. Codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Residual standard error:	**1.305 on 20 degrees of freedom**
Multiple R-squared:	0.9992,
Adjusted R-squared:	0.9988
F-statistic:	2755 on 9 and 20 DF,
p-value:	< 2.2e-16

The model estimated 10 parameters from the data (10 rows in the R output): five intercepts and five slopes. The first day (day was the unit used as factor) is used as a baseline against which to compare the other four days.

The coefficients -0.04917 (*Intercept*) and 1.00558 (*Std*) represent the intercept and the slope of the regression line for day 1. For the second day, the intercept and the slope are given by the sum, respectively, of the first and second quantities (-0.04917 + -0.56464 = -0.61381) and the sum of the sixth and seventh quantities (1.00558+0.02242 =1.028). The other days' regression parameters can be computed in the same way by summing the proper rows.

The last column on the right indicates the parameter values which are significantly different from zero when compared with day 1. The table shows that intercepts (first five rows) and slopes (last five rows) do not differ significantly from day 1 at a level of significance of 5%. However, this model compares, using a t-test, the slopes and the intercepts among days only with the slope and intercept for day 1.

To test the hypothesis of equal slopes of regression lines for several different days, the complete model obtained with the interaction must be compared with the model for which the parallelism hypothesis is considered valid. The model with equal slope is given by:

```
model_2 <- lm(Computed ~ Day + Std, Edata)
```

and the comparison is obtained with the R code:

```
anova(model_1, model_2)
```

The output of the ANOVA command is:

Analysis of Variance Table

Model 1:		Computed ~ Day * Std				
Model 2:		Computed ~ Day + Std				
Res.	Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	20	34.067				
2	24	44.294	-4	-10.27	1.501	0.2397

From the output, the p-value from the F test is higher than 0.05 (Pr=0.2397) and the null hypothesis of equal slopes between the 5 regression lines is therefore accepted. At this point it is possible to test the equality of the intercepts. This is done by comparing the previous model (equal slopes) with the model which assumes equal regression lines (equal slopes and equal intercepts).

```
Model_3 <- lm(Computed ~ Std, Edata)
anova(model_2, model_3)
```

The output is:

Analysis of Variance Table

Model 1:		Computed ~ Day * Std				
Model 2:		Computed ~ Day + Std				
Res.	Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	24	44.294				
2	28	44.308	-4	-0.0135	0.0018	1

Based on the results, the hypothesis of equals regression lines (Pr>0.05) is accepted, which implies that the day on which the calibration is computed does not influence the output variable (concentration of the analyte).

The same results can be obtained by another R command which results in an ANOVA table with the summary parameters:

```
> model_B<-aov(Computed~Day*Std, Edata)
> summary(model_B)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Day	4	0	0	0.002	1
Std.	1	42226	42226	24789.509	<2e-16***
Day:Std.	4	10	3	1.501	0.24
Residual	20	34	2		
Signif. Codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

The F values and the corresponding probability values for the interaction term (Day:Std) and for the intercept (Day) are the same, as found in the previous computations taken separately. Again, this indicates that there is no significant difference between the slopes and the intercepts of the calibration curves, at a level of confidence of 95%.

4.2. Estrone: case b. – Surface water

It is easier to apply the ANCOVA model to case b, which has only two different slopes and intercepts to compare. The model is specified using the following formula:

```
> model_1<-lm(Computed~Day*Std, Edata)
> summary(model)
```

The output of the ANCOVA is the following:

Call:

lm(formula = Computed ~ Day * Analyte, data = Edata)

Residuals:				
Min	1Q	Median	3Q	Max
-0.26151	-0.09115	0.01426	0.11512	0.31967

Coefficients:				
	Estimate	Std. Error	T value	Pr(>|t|)
(Intercept)	0.055914	0.101366	0.552	0.5963
DayDay2	-0.162130	0.153353	-1.131	0.2908
Analyte	1.000111	0.002210	452.496	<2e-16***
DayDay2:Analyte	0.006439	0.003126	2.060	
Signif. Codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Residual standard error:	**0.2015 on 8 degrees of freedom**
Multiple R-squared:	1
Adjusted R-squared:	1
F-statistic:	1.374e+05 on 3 and 8 DF,
p-value:	< 2.2e-16

The model estimated four parameters from the data (four rows): two intercepts and two slopes. With only two groups to compare (day 1 and day 2), the t-test is sufficient to verify the hypothesis of equal slopes.

From the results, the hypothesis of equal slope between the regression lines is accepted (p-level = 0.0733), with a level of confidence of 95%.

At this point, it is possible to adopt an additive model which determines the two regression lines with the same slope (parallel lines). From this model, it is possible to test the hypothesis of equal intercepts:

> model_2 <- lm(Computed ~ Day + Std, Edata)

```
> summary(model_2)
```

The output is:

```
Call:
lm(formula = Computed ~ Day + Std, data = Edata)
```

Residuals:				
Min	1Q	Median	3Q	Max
-0.33685	-0.11951	0.02812	0.03314	0.37376

Coefficients:				
	Estimate	Std. Error	T value	Pr(>|t|)
(Intercept)	-0.030376	0.107667	-0.282	0.784
DayDay2	0.010450	0.135687	0.077	0.940
Std				
Signif. Codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Residual standard error:	**0.235 on 9 degrees of freedom**
Multiple R-squared:	1
Adjusted R-squared:	1
F-statistic:	1.514e+05 on 2 and 9 DF,
p-value:	< 2.2e-16

From the output, the p-value for the coefficient Day2 is 0.940 and therefore the null hypothesis of equal intercepts between the two regression lines is accepted.

Finally, the two calibration curves derived from the analysis of Estrone in surface water in two different days can be assumed coincident at a level of confidence of 95%.

The same results can be obtained by a summary R command which results in an ANOVA table with the following parameters.

```
> model_B<-aov(Computed~Std*Day, Edata)
> summary(model_B)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Std.	1	16732	16732	4.121e-05	<2e-16***
Day	1	0	0	8.000e-3	0.9306
Std:Day	1	0	0	4.244e+0	0.0733
Residual	8	0	0		
Signif. Codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

As in the case a, the F and probability values for the interaction term (Day x Std) and for the intercept (Std) are higher than 0.5, indicating that the hypothesis of equal slopes and intercepts is satisfied.

4.3. Estrone: case c. – Surface vs. MilliQ water

After separately testing the day-to-day comparability of calibration curves in the MilliQ and Surface waters, it is possible to compare the calibration curves between the two matrices. In this case, the ANCOVA will give us information on the effect of the matrix type. The calibration curve of the first day for each matrix type was taken for performing the ANCOVA.

Since case c is similar to case b, only the summary R command is given.

```
> model<-aov(Computed~Std*Day, Edata)
> summary(model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Std.	1	16716	16716	2.189e+05	<2e-16
Day	1	0	0	6.8000e-02	0.802
Std:Day	1	0	0	1.627e+00	0.238
Residual	8	1	0		
Signif. Codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Based on the results, the hypothesis of equal slope between the regression lines is accepted (p-level of the interaction term = 0.238) at a 95% level of confidence. Moreover, the p-values for the effect of the Day parameter is 0.802, indicating equal intercepts between calibration curves.

Finally, the two calibration curves derived from the analysis of Estrone in surface and MilliQ waters can be assumed to be coincident at a level of confidence of 95%. This implies that there is no a significant effect of matrix type on calibration curves for the considered analyte.

The procedure described for the Estrone compound was also applied to the 17β-Estradiol compound. Only the summary results are given here.

4.4. 17β-Estradiol case a – MilliQ water

For case a, the ANOVA summary table is the following:

```
> model_B<-aov(Computed~Day*Std, Edata)
> summary(model_B)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Std.	1	51515	51515	61977.448	<2e-16***
Day	4	0	0	0.001	1.000
Std:Day	4	4	1	1.113	0.378
Residual	20	17	1		
Signif. Codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

The probability for the interaction term (Day:Std) is equal to 0.378, and the hypothesis of equal slopes between the five calibration curves is accepted.

In the case of the intercept (Day parameter), the p-value is 1.0, and even intercepts are comparable. Again, this indicates that there is no significant difference between the slopes and the intercepts of the calibration curves, at a level of confidence of 95%.

4.5. 17β-estradiol case b – Surface water

From the summary computation, the R output is:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Std.	1	20626	20626	3.12e+04	1.18e-15***
Day	1	0	0	3.00e-2	0.867
Std:Day	1	0	0	6.00e-03	0.942

Residual	8	5	1		
Signif. Codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				

For the interaction term (Day x Std), the probability value P higher is 0.942, indicating that there is no significant difference between the slopes of the calibration curves. Moreover, the p-values for the effect of the Day parameter is 0.867, indicating equal intercepts between calibration curves.

4.6. 17β-estradiol case c – Surface vs. MilliQ water

After separately testing the daily comparability of calibration curves in the MilliQ an d Surface waters, it is possible to compare the calibration curves between the two matrices. The output result is:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Std.	1	20528	20528	21443.715	5.29e-15***
Day	1	0	0	0.006	0.939
Std:Day	1	0	0	0.169	0.692
Residual	8	8	1		
Signif. Codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

For the interaction term (Day x Std), the probability value P higher is 0.692, indicating that there is no significant difference between the slopes of the calibration curves. Moreover, the p-value for the effect of the Day parameter is 0.939, indicating equal intercepts between calibration curves. This implies that the matrix type has no significant effect on calibration curves for the considered analyte.

5. Conclusions

In conclusion, for Estrone and 17β-Estradiol, the calibration curves determined in MilliQ and in surface waters are coincident (same slopes and same intercepts). For method validation purposes, the absence of a matrix effect over the concentration range of interest means than no new method validation needs to be carried out when the matrix type changes.

Finally, the method validated for Estrone and 17β-Estradiol in MilliQ water can also be used for their determination in surface waters.

LB-NA-27970-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub
EU Science Hub - Joint Research Centre
Joint Research Centre
EU Science Hub

Publications Office

doi:10.2788/122189
ISBN: 978-92-79-59357-4